

A predictive model for the determination of cadmium concentration in cocoa beans using laser-induced plasma spectroscopy

Modelo predictivo para la determinación de la concentración de cadmio en granos de cacao mediante espectroscopia de plasma inducido por láser

Sandra Liliana Herrera Celis¹, Jáder Enrique Guerrero Bermúdez^{2*}, Enrique Mejía-Ospino¹, and Rafael Cabanzo Hernández¹

ABSTRACT

This study proposes a predictive model to determine the concentration of cadmium (Cd) in cocoa beans based on laser-induced breakdown spectroscopy (LIBS) and partial least squares regression (PLSR-1 or PLS-1). The multivariate calibration model was developed using 46 cocoa bean samples, with Cd concentrations up to 1 mg kg⁻¹. The increase of the LIBS signal in the Cd emission lines was evident when the cocoa bean sample was subjected to a solid-liquid-solid transformation (SLST). The range error ratio (RER) was 7.92, which allowed it to be classified as a screening model. Monte Carlo cross-validation was used, with 60% of samples for calibration and the remaining for testing. The standard error of cross-validation (SECV) and standard error of calibration (SEC) were 0.12 mg kg⁻¹ and 0.05 mg kg⁻¹, respectively. The proposed procedure is framed within the alternatives for the chemical analysis of cocoa.

Key words: inorganic contaminants, heavy metals, partial least square regression, atomic spectroscopy.

RESUMEN

Este estudio propone un modelo para predecir la concentración de cadmio (Cd) en granos de cacao basado en espectroscopia de plasma inducido por láser (LIBS) y regresión por mínimos cuadrados parciales (PLSR-1 o PLS-1). El modelo de calibración fue desarrollado a partir de 46 muestras de granos de cacao con concentración no mayor a 1 mg kg⁻¹. El incremento en la señal LIBS fue evidente cuando la muestra de grano de cacao fue sometida a una transformación sólido-líquido-sólido (SLST). La razón del rango de error (RER) es 7.92, lo que permite determinar que el modelo es de tamizaje. Se utilizó la estrategia de validación cruzada Montecarlo con el 60% de las muestras para calibración y las restantes para prueba. El error estándar de validación cruzada (SECV) y de calibración (SEC) fue 0.12 mg kg⁻¹ y 0.05 mg kg⁻¹, respectivamente. El procedimiento propuesto se ubica en el marco de las alternativas de inspección y análisis químico del cacao.

Palabras clave: contaminantes inorgánicos, metales pesados, regresión por mínimos cuadrados parciales, espectroscopia atómica.

Introduction

Cocoa (*Theobroma cacao* L.) and its derivatives can be considered commodities with health benefits due to their high content of polyphenols and antioxidants (Oliveira *et al.*, 2021). Nevertheless, cocoa has been identified as an important source of cadmium (Cd), which is a transition metal without biological function and of recognized toxicity in humans that affects the kidneys and calcium absorption in bones (Järup & Akesson, 2009; Satarug, 2018).

Cadmium is released naturally in soils through the weathering of rocks, where it is usually found in concentrations of less than 0.2 mg kg⁻¹ (Gramlich *et al.*, 2017). However, anthropogenic activities such as mining, agrochemical

industry (fertilizers, pesticides), in addition to its considerable mobility in soil, significantly increase its content. Cadmium absorption by cocoa plants occurs through the roots and is conducted through its vascular system to finally reside in the leaves and fruits (Checa *et al.*, 2019; Vanderschueren *et al.*, 2020).

The raw material for the cocoa industry comes mainly from Africa, Central and South America. For the latter, numerous studies have shown that cocoa beans (nibs and shells) have relatively high concentrations of cadmium (Bertoldi *et al.*, 2016; Argüello *et al.*, 2019; Rodríguez Albarracín *et al.*, 2019; Bravo *et al.*, 2021). In the Americas, Colombia has the potential as a bean exporter with increasing production. In the period 2020-2021, production reached 70205 t

Received for publication: September 22, 2022. Accepted for publication: November 30, 2022

Doi: 10.15446/agron.colomb.v40n3.104911

¹ Laboratorio de Espectroscopia Atómica y Molecular (LEAM), Centro de Materiales y Nanociencias (CMN), Universidad Industrial de Santander, Parque Tecnológico Guatiguará, Piedecuesta, Santander (Colombia).

² Grupo de Óptica y Tratamiento de Señales, Universidad Industrial de Santander, Bucaramanga, Santander (Colombia).

* Corresponding author: jader@uis.edu.co



of dry beans, corresponding to its highest historical record (Fedecacao, 2021). Around 95% of exported beans were categorized as Fine Flavour Cocoa (FFC). FFC is a market category established by the International Cocoa Organization (ICCO) that recognizes sensory attributes such as flowery, fruity, caramelly, and nutty (Escobar *et al.*, 2020; Escobar *et al.*, 2021).

In general, atomic emission/absorption spectral techniques allow for the quantitative determination of Cd concentrations in cocoa beans, among them: inductively coupled plasma mass spectrometry (ICP-MS), inductively coupled plasma optical emission spectrometry (ICP-OES), graphite furnace atomic absorption spectrometry (GF-AAS), and flame atomic absorption spectrometry (FAAS). These methods for the assessment of Cd concentrations in cocoa are characterized by their robustness and limit of detection (LOD). In ICP-OES, for example, we found LODs near 0.043 mg kg^{-1} (Rodríguez *et al.*, 2022). However, the large number of supplies, as well as the generation of environmentally aggressive waste, lead to a search for analysis strategies that minimize these factors. By contrast, laser-induced breakdown spectroscopy (LIBS) has become a useful technique for food analysis that is relatively simple and with a minimum production of polluting waste.

LIBS is suitable for the detection of cadmium in agricultural materials and food (Menegatti *et al.*, 2019; Nicolodelli *et al.*, 2019; Senesi *et al.*, 2019). Chemometrics and LIBS have been proposed for the determination of Cd in fruits and vegetables (Yao *et al.*, 2017; Shen *et al.*, 2018). Zhao *et al.* (2019) determine Cd in lettuce leaves using enhanced LIBS by adding silver nanoparticles (NELIBS); the enhanced signal of the 214.4 nm line allows prediction of Cd concentrations less than $60 \text{ } \mu\text{g kg}^{-1}$ (Zhao *et al.*, 2019). Yang *et al.* (2019) proposes a simple and low-cost pre-treatment method for rice grain samples, called solid-liquid-solid transformation that improves the signal to noise ratio for Cd lines in LIBS spectra, obtaining detection limits near $2.8 \text{ } \mu\text{g kg}^{-1}$. Very recently, univariate and multivariate calibration methods, such as partial least squares regression (PLSR), least squares support vector machines (LS-SVM), and extreme learning machines (ELM) in combination with LIBS were proposed for quantitative analysis of Cd in the rice roots (Wang *et al.*, 2021).

Previous studies on the determination of Cd in cocoa beans using ICP-OES were developed by Chavez *et al.* (2015) and Oliveira *et al.* (2021). Gramlich *et al.* (2017) study cadmium uptake in soils and cocoa beans by FAAS. Also, in Colombia, the organs of cocoa plants and soil are analysed by ICP-MS

for the determination of heavy metals, with emphasis on Cd (Aguirre-Forero *et al.*, 2020).

In this research, a predictive model was proposed to determine Cd concentration in cocoa beans of Colombian origin, based on LIBS, FAAS (as a reference technique) combined with PLSR-1 (partial least squares regression). This took advantage of the agility to generate LIBS spectra that together with the Cd concentrations determined by FAAS and the chemometric procedures allowed the construction of the model and a reduction in the time of analysis with respect to conventional spectral methodologies. If the procedure proposed in this study is compared with the FAAS technique, the time required to obtain LIBS data is estimated as 1 h (including SLST transformation). In contrast, FAAS requires digestion processes that typically take more than 48 h.

Materials and methods

Samples of cocoa beans of Colombian origin that were dried and ground were collected for one year from different laboratories. Prior to calibration, all samples were homogenized using a mortar.

Determination of cadmium concentration by FAAS: reference method

Sample treatment

Initially 65 samples of cocoa beans of Colombian origin were treated for analysis of the concentration of cadmium by FAAS. The cocoa beans were obtained from different plantations, but the precise geographic location is unknown.

Prior to microwave oven assisted digestion (MWAD), 0.5 g of cocoa beans was immersed in 8 ml of HNO_3 (MERCK, 65%), for 48 h, using teflon vessels (adapted from Oliveira *et al.* (2021)). After this period, 2 ml of H_2O_2 (30%) was added. Samples were introduced into the microwave digestion system (Multiwave GO Anton Paar, GmbH), according to the following procedure: (a) starting from room temperature to reach 80°C in 5 min, remaining at that temperature for 5 min; (b) from 80°C to 150°C for 5 min and staying at that temperature for 5 min; (c) from 150°C to 180°C for 3 min and remaining at that temperature for 25 min. The teflon vessels were allowed to cool to room temperature and carefully opened. The digestion product was transferred to a 25 ml balloon and filled with HNO_3 (0.5%). The washing of the glass and teflon material was carried according to AOAC Official (Jorhem & Engman, 2000).

Atomic absorption analysis

The standard solution was obtained by dissolving 1 g of Cd powder (Cadmium EMSURE[®] Merck KGaA, Darmstadt, Germany, particle size 0.3-1.6 mm) in 21 ml of HNO₃, in a 1:1 ratio (v/v) and then filled to 1 L with deionised water. From this standard solution a calibration curve was constructed with six Cd concentrations: 0.02, 0.5, 1, 1.5, 2, and 3 mg kg⁻¹.

The analysis was performed on a FAAS spectrometer (Thermo Electron Corporation, USA). The wavelength of the radiation emitted by the Cd hollow cathode lamp was 228.8 nm (current 4 mA). The flame was an air/acetylene oxidising flame with a flow rate of 1.2 L min⁻¹ and a flame height of 7 mm. Slit width was 0.5 mm. Each concentration was determined by triplicate and its average value (\pm standard deviation) is shown in Table 1.

TABLE 1. Cadmium concentrations (mean \pm standard deviation (SD)) determined by FAAS in 46 samples of dry cocoa beans. In the calibration of the predictive model, only concentrations below 1 mg kg⁻¹ were considered (negative concentrations that resulted from the FAAS curve were discarded).

Sample	Cd concentration Mean \pm SD (mg kg ⁻¹)	Sample	Cd concentration Mean \pm SD (mg kg ⁻¹)
M01	0.29 \pm 0.01	M24	0.02 \pm 0.01
M02	0.84 \pm 0.06	M25	0.05 \pm 0.01
M03	0.72 \pm 0.04	M26	0.05 \pm 0.01
M04	0.32 \pm 0.01	M27	0.02 \pm 0.01
M05	0.23 \pm 0.03	M28	0.02 \pm 0.01
M06	0.61 \pm 0.04	M29	0.30 \pm 0.01
M07	0.96 \pm 0.10	M30	0.03 \pm 0.01
M08	0.54 \pm 0.01	M31	0.03 \pm 0.01
M09	0.61 \pm 0.03	M32	0.29 \pm 0.01
M10	0.19 \pm 0.01	M33	0.05 \pm 0.02
M11	0.32 \pm 0.01	M34	0.05 \pm 0.01
M12	0.22 \pm 0.01	M35	0.03 \pm 0.01
M13	0.17 \pm 0.04	M36	0.08 \pm 0.02
M14	0.17 \pm 0.04	M37	0.02 \pm 0.01
M15	0.17 \pm 0.01	M38	0.02 \pm 0.01
M16	0.17 \pm 0.01	M39	0.0058 \pm 0.0015
M17	0.16 \pm 0.01	M40	0.07 \pm 0.01
M18	0.13 \pm 0.01	M41	0.21 \pm 0.01
M19	0.17 \pm 0.01	M42	0.15 \pm 0.02
M20	0.29 \pm 0.46	M43	0.16 \pm 0.01
M21	0.06 \pm 0.01	M44	0.28 \pm 0.02
M22	0.06 \pm 0.03	M45	0.96 \pm 0.01
M23	0.03 \pm 0.01	M46	0.06 \pm 0.01

Detection of cadmium in cocoa beans by LIBS

Determination of cadmium in pellets

In LIBS, the most common process for the analysis of samples is the formation of pellets after its homogenization (Sezer *et al.*, 2017; Yang *et al.*, 2018; Senesi *et al.*, 2019). Pellets were produced from 500 mg of cocoa beans, previously macerated, by applying 2 t cm⁻² of pressure for 2 min using a hydraulic press (Carver Inc., USA). However, it was only possible to determine Cd in pellets previously contaminated with a high concentration (>15 mg kg⁻¹) of this element. To detect Cd in samples with lower concentrations, it was necessary to implement the SLS transformation, proposed by Yang *et al.* (2019) that is described in the next section. SLST is a commonly used procedure for the quantification of heavy metals in food samples. In cocoa beans, Cd is bound in the chemical forms with other elements (among others: CdCl₂, CdOH⁺, CdCl₃⁻ and organic ones). At low concentrations of Cd, it is difficult to extract it by laser. For this reason, the addition of 0.1 N HCl allows the release of Cd²⁺ and the formation of soluble salts with the Cl⁻ ion (Yang *et al.*, 2019).

Solid-Liquid-Solid Transformation (SLST)

To 1 g of macerated cocoa beans, 14 ml of 0.1 N HCl solution was added. The solution was submitted to an ultrasonic bath for 15 min. From the supernatant, 200 μ l were taken, deposited on a glass sample holder, and allowed to dry at a temperature of 70°C. In this way, a layer of the material was obtained; the process was repeated to a total of five layers (Yang *et al.*, 2019).

The experimental parameters were as follow: added volume of 0.1 N HCl solution (10, 12, 14, and 16 ml), and the following ultrasonic bath times using a Branson ultrasonic cleaner 1510, Japan of 5, 10, 15, and 20 min were studied according to the criterion of maximising the area under the spectral curve of the (concatenated) array of the 214.44 nm and 226.50 nm emission lines of the LIBS spectra. Unlike the volume of solution added and the ultrasonic time, the number of layers did not exhibit a maximum (the area under emission line). On the contrary, from the second layer, it grew monotonically as the number of layers increased. Given that more layers mean longer sample preparation time, five layers resulted in a good compromise between sample preparation time and signal intensity.

The emission line of Cd 214.44 nm, Figure 1A-C depicted the behaviour of these experimental parameters. Similar characteristics followed for the Cd 226.50 nm emission line. Table 2 shows the experimental parameters used for the SLS transformation.

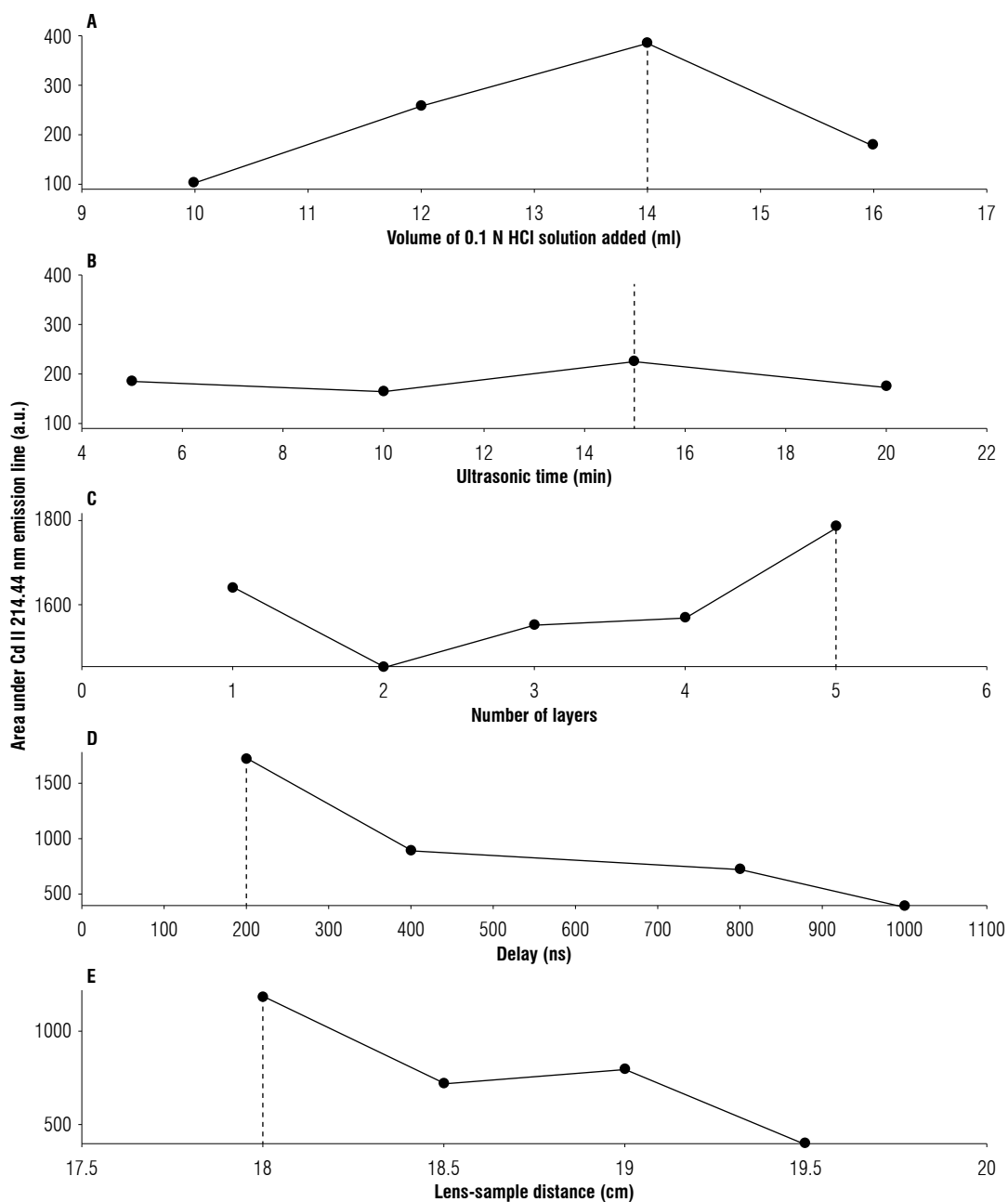


FIGURE 1. Exploration of experimental and instrumental parameters used in the solid-liquid-solid transformation (SLST) and LIBS setup: area under emission lines (214.44 nm) in arbitrary unit (a.u.) vs. (A) Volume of HCl 0.1 N solution added, (B) ultrasonic time, (C) number of layers, (D) delay, and (E) lens-sample distance.

TABLE 2. Result of the optimization of the parameters of the solid-liquid-solid transformation.

SLST-Parameter	Value
Volume of 0.1N HCl solution added	14 ml
Ultrasonic time	15 min
Number of layers	5

Optimization of instrumental parameters for LIBS

LIBS spectra were recorded using a Q-switched Nd:Yag laser (Q-smart, Quantel, Inc., USA) operating at 1064 nm.

Adopting the same optimization criterion as in the SLST, we proceeded to set the LIBS instrumental parameters, namely, the delay and lens-sample distance. For a five-layer sample, delays of 200, 400, 800, and 1000 ns were tested with a lens-sample distance of 19.5 cm and 100 shots. Keeping the delay and number of shots constant (200 ns and 100 shots, respectively), the lens-sample distance was varied in the range 18 cm to 19.5 cm. Finally, the number of laser shots on the sample was between 10 and 200, at delay (200 ns) and lens-sample distance (19.5 cm). To

ensure that each laser impact covered a different region on the sample, the glass sample holder was supported on a turntable (with stepper motor). The selected parameters were as follow: a delay of 200 ns, integration time of 5 μ s, lens-sample distance of 18 cm and number of shots of 200, pulse duration 5 ns, a repetition rate 10 Hz. As the number of shots increased so did the area under the emission line; however, the sample size limited this number to 200. The laser energy was 150 mJ/pulse. The behaviour of the area under the emission line 214.44 nm against the delay time and the sample lens distance is shown in Figure 1 D-E.

The plasma emission was focused through a quartz lens (focal length, 50 mm) and collected with a quartz optical fiber and introduced into a 0.5 m Czerny-Turner spectrograph (Shamrock 500i, Andor Technology, USA). A two thousand and four hundred grooves/mm diffraction grating (spectral resolution 0.075 nm, dispersion 0.83 nm mm⁻¹, and blazed 250 nm) was used to disperse the emission spectrum that was projected on the image plane of the spectrograph and recorded on an intensified charge-coupled device (ICCD) detector with an array of 1024 \times 256 pixels (iStar DH720, Andor Technology, USA). The characteristics of the LIBS spectra acquisition setup are shown in Table 3.

TABLE 3. Characteristics of the LIBS spectra acquisition setup.

Instrumental parameter for LIBS	Value
Laser wavelength	1064 nm
Delay time	200 ns
Number of laser shots	200
Lens-sample distance	18 cm
Laser energy/pulse	150 mJ
Integration time	5 μ s
Pulse duration	5 ns
Repetition rate	10 Hz

Figure 2 shows the LIBS signal of cocoa bean sample 21 (0.06 ± 0.01 mg kg⁻¹) before and after SLS transformation. The increase of the signal in the Cd II emission line 214.44 nm is evident.

PLSR model development

PLSR-1 is a well-known chemometric technique for the implementation of predictive models (Otto, 2007). It is characterized by its simplicity and the possibility of obtaining good performance without over-fitting. In developing a prediction model, a compromise must be maintained between the number of latent variables (predictors) and the degrees of freedom (ASTM E1655-05, 2012).

In a PLS-1 predictive model, the relationship between the predictors \mathbf{x} , and the estimated response, \hat{y} , is established according to Equation 1:

$$\hat{y} = b\mathbf{x} \quad (1)$$

where \hat{y} corresponds to an estimated concentration of Cd in cocoa beans; b is the vector of coefficients of the regression in the partial least squares sense (including the independent term); and \mathbf{x} is a vector consisting of 33 intensity values (concatenated, with baseline correction and normalised by the area under the spectral curve) of the emission lines 214.44 nm and 226.50 nm. The concatenated array is formed by uniformly sampling over 17 and 16 points the intervals [214.3, 214.7] and [226.4, 226.7] (Fig. 3). Although in FAAS the 228.88 nm emission line is the reference for the measurements, it was not considered for the development of the predictive model due to interference with the iron emission lines.

The multivariate calibration was performed with Monte Carlo cross-validation (MCCV) to ascertain the complexity

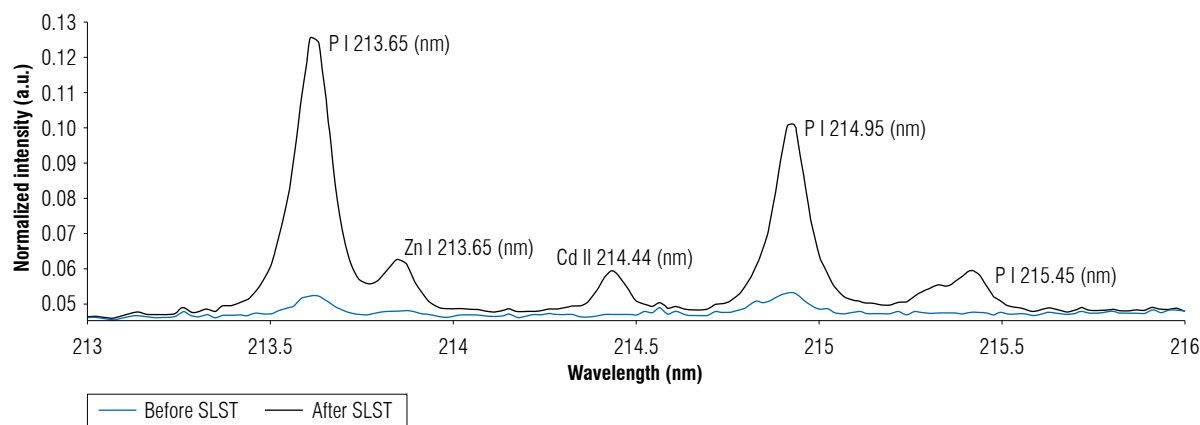


FIGURE 2. The increase of the LIBS signal in the Cd II 214.44 nm spectral line is evident when the cocoa bean sample 21 (0.06 ± 0.01 mg kg⁻¹) is subjected to a solid-liquid-solid transformation (SLST). Normalized intensity in arbitrary unit (a.u.).

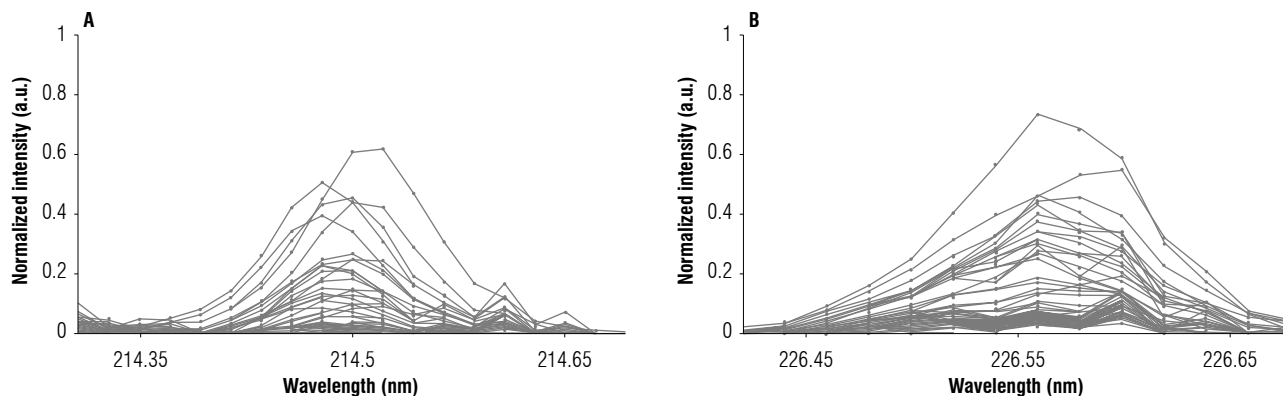


FIGURE 3. LIBS signal normalized by the area under the spectral curve of the Cd II 214.44 nm (A) and Cd II 226.50 nm (B) emission lines. Thirty-three (33) equally spaced points are concatenated to form an array that constitutes the input signal in the predictive model. Normalized intensity in arbitrary unit (a.u.).

of the model, *i.e.*, the number of latent variables (LVs) of the PLS model (Xu & Liang, 2001). This cross-validation strategy requires splitting the samples into two sets, one with n_c samples for training (testing or calibration), and the other with n_{cv} for validation. In all cases, the training set contained the samples with the highest and lowest Cd concentration values. Therefore, the regression model is only valid for interpolating values in that range.

The criteria used to define the number of LVs were the following: the square root of the standard error of calibration (SEC) and cross-validation (SECV), defined according to Equations 2 and 3:

$$SEC = \sqrt{\frac{1}{d_c} \sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2} \quad (2)$$

$$SECV = \sqrt{\frac{1}{d_{cv}} \sum_{i=1}^{n_{cv}} (\hat{y}_i - y_i)^2} \quad (3)$$

where $d_x = n_x - (k+1)$ (with x , c or cv for calibration or cross-validation samples, respectively) corresponds to the degrees of freedom, and k is the number of latent variables (LVs). In Equations 2 and 3, \hat{y}_i is the estimated value of Cd concentration corresponding to the i -th sample with a measured value y_i . Another criterion used to determine the number of LVs was the correlation between two successive \bar{b} -vectors (the \bar{b} -vectors are the average of the b -vectors resulting from each partition, according to Equation 1) (Andrade-Garda, 2009).

Other figures of merit, which allowed us to evaluate the performance of the calibration model developed were as follows: the F-test with degrees of freedom, $(k-1)$ (numerator) and $(n_c - k)$ (denominator) given by Equation 4:

$$F(k-1, n_c - k) = \frac{\sum_{i=1}^{n_c} (\hat{y}_i - \bar{y})^2 / (k-1)}{\sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2 / (n_c - k)} \quad (4)$$

where \bar{y} is the average of the Cd concentrations used in the calibration group.

The range error ratio, RER, defined according to Equation 5, was calculated as follows:

$$RER = \frac{|y_{MAX} - y_{MIN}|}{SECV} \quad (5)$$

with y_{MAX} and y_{MIN} being the maximum and minimum values of Cd concentration in cocoa beans (range for cross-validation sample set), respectively, determined by FAAS.

The confidence limit at a level of 95% for the performance of the model was calculated as $\hat{y}_i \pm t \cdot SEC \cdot \sqrt{(1 + h_{ii})}$, where t is the student's t value for d_c degrees of freedom. For each training sample, with an estimated value \hat{y}_i , h_{ii} is a scalar taken in order from the main diagonal of the matrix $T \times T'$, where scores T are the coordinates of the sample in the principal components space (ASTM E1655-05, 2012). Finally, as usual, the coefficients of determination for both the set of calibration, R_c^2 and cross-validation R_{cv}^2 are reported.

Results and discussion

The limit of detection (LOD) for FAAS analysis resulted in 0.056 mg kg^{-1} . This figure of merit was calculated according to the expression: $LOD = C_{blank} + 3\sigma$, where C_{blank} corresponds to an average concentration of ten blanks with $\sigma = 0.0019 \text{ mg kg}^{-1}$ (standard deviation). Except for the addition of cocoa, the blanks were subjected to the treatment previously described in the sample treatment section. A recovery percentage of 98% was obtained, which means that the sample treatment process and the analytical measurement is acceptable. From the initial set of 65 samples, those whose FAAS analysis resulted in a negative concentration were

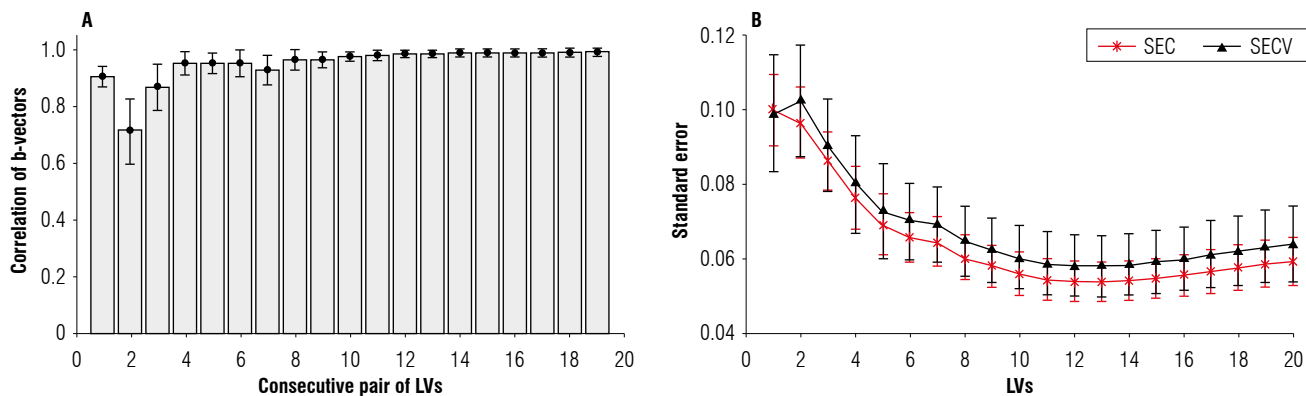


FIGURE 4. Criteria for selecting the dimension (number of latent variables, LVs) of the calibration model. A) corresponds to the correlation of two consecutive \bar{b} -vectors; B) the standard error of calibration (SEC) and standard error of cross-validation (SECV) characteristic versus the number of LVs. In A) and B), vertical bars indicate the standard deviation from the average of 1000 partitions.

discarded (negative value means, in this case, Cd concentration below of limit of detection). In addition, due to the small number of samples with concentrations above 1 mg kg^{-1} , the predictive model was limited to this value, taking 46 samples for its development.

The concentration of Cd in cocoa beans using LIBS and FAAS was carried out by applying PLS-1 implemented on Matlab R2019b (MathWorks, Inc.) using the statistically inspired modification of the partial least-square (SIMPLS) algorithm (De Jong, 1993; Otto, 2007; Faber & Ferré, 2008). A total of 1000 partitions generated the same number of models for each of the first 20 LVs. Each partition consisted of 60% of the samples for training (28 samples) and the remaining 40% for validation (18 samples). Usually, the number of LVs in the model is selected according to a minimum in SECV (Fig. 4B) that is suggested in this case to be between 12 and 13 LVs.

To respect the degrees of freedom of the calibration mode, this quantity of LVs requires a larger number of samples than those available. For this reason, we use the criterion given by the correlation of two consecutive \bar{b} vectors. In Figure 4A, typical oscillations in this correlation are observed up to the eighth latent variable. Beyond this number of LVs, the correlation tends to remain constant, indicating that there is scarce new information about the model. Based on this behaviour, we selected eight LVs for the construction of the predictive model. The uncertainty reported (vertical bars) were the standard deviations, for each metric, from the 1000 partitions using \bar{b} -vectors.

A moderately optimistic model could be proposed by selecting a partition with SECV greater than (or equal to) SEC, and with the smallest distance to the line of identity

in the SECV-SEC plane (Fig. 5). It is an empirical fact that when SECV and SEC are calculated for different partitions, using \bar{b} -vector in Equation 1, a set of points distributed over an arc segment is obtained. In the SECV-SEC space, moderately optimistic models are located close to the identity line (Niño *et al.*, 2019). Thus, taking the partition closest to this line, the performance metrics of the proposed model are calculated.

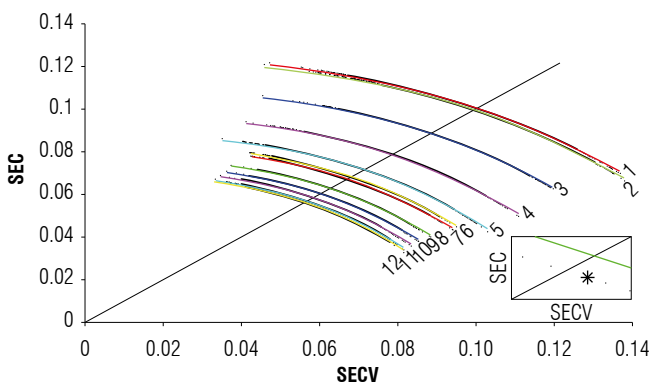


FIGURE 5. In the standard error of cross-validation (SECV)-standard error of calibration (SEC) plane, averages over 1000 partitions are placed on arc segments. For the number of latent variables defined, a partition is selected that satisfies the condition SECV is greater than or equal to SEC. With this partition, the performance metrics for the predictive model are estimated.

For the selected partition, in the eighth LV, the SEC has a value of $0.05 \pm 0.0066 \text{ mg kg}^{-1}$ with the SECV $0.12 \pm 0.0092 \text{ mg kg}^{-1}$. Similarly, the coefficient of determination for calibration and cross-validation samples, R_C^2 and R_{CV}^2 are 0.97 ± 0.0093 and 0.79 ± 0.0042 .

The F_{Cal} (calculated F) was 99.23, which must be compared with the $F_{Tab}(0.95, 7, 20)$, (tabulated F , with a confidence

margin of 95%, $N_C=28$ and $k=8$). Since $F_{Cal} > F_{Tab}=3.01$, we can argue that the model appropriately fits the data. On the other hand, the selected partition has a RER of 7.92 ± 2.92 , therefore the proposed model can be considered for screening. The performance metrics for the proposed model are summarised in Table 4.

TABLE 4. Parameters describing the performance of the proposed calibration model.

Parameter	Value
Cd concentration range	[0.01-1] mg kg ⁻¹
LVs	8
R_C^2	0.97 ± 0.0093
R_{CV}^2	0.79 ± 0.0042
RER	7.92 ± 2.92
SEC	0.05 ± 0.0066 mg kg ⁻¹
SECV	0.12 ± 0.0092 mg kg ⁻¹
F_{Cal}	99.23
F_{Tab} (0.95,7,20)	3.01
β_1 (p-value)	0.50
β_2 (p-value)	0.34

LVs, latent variables; R_C^2 and R_{CV}^2 are the coefficient of determination for calibration and cross-validation; range error ratio (RER); SEC, standard error of calibration; SECV, standard error of cross-validation; F_{Cal} and F_{Tab} are the F-test statistic values, calculated and tabulated, respectively; p-values for linear (β_1) and quadratic (β_2) coefficient distributions.

Figure 6 depicts the performance of the proposed calibration model. Note that a small fraction of the samples falls outside the 95% confidence bands (dashed lines). For more details, this information (including residuals, written in

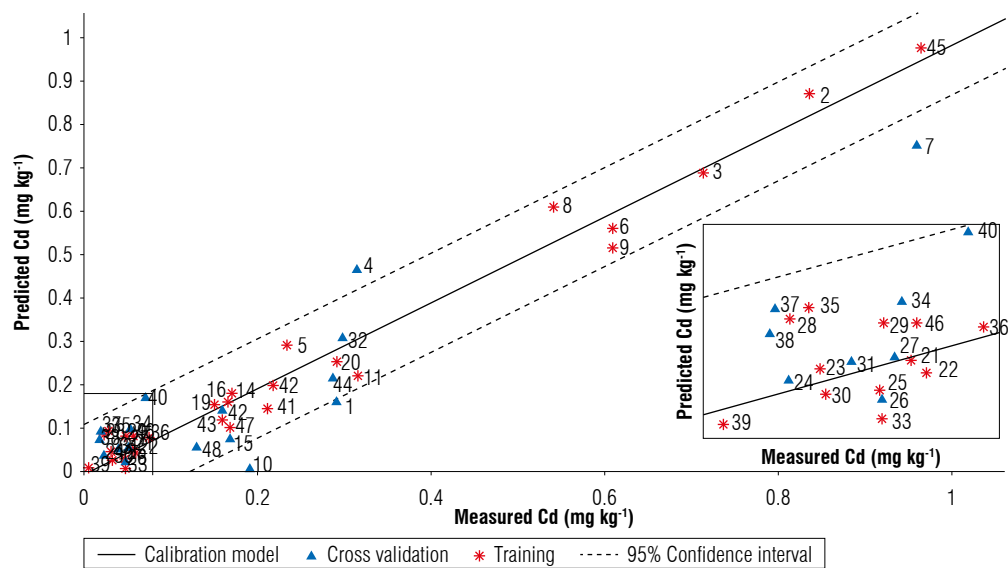


FIGURE 6. Correlation of measured values of Cd concentrations and values predicted by the PLSR-1 model. Dashed lines constitute the 95% confidence band. Samples with measured concentrations close to 0.1 mg kg⁻¹ are presented in the box.

brackets) is shown in Table 5. The correlation shows that the model responds to the performance indicators explained in the previous paragraph, following the procedure suggested for the evaluation of trends in residuals of multivariate calibration models by permutation test (Filgueiras *et al.*, 2014). A total of 50,000 permutations allows us to obtain distributions of the linear (β_1) and quadratic (β_2) coefficients, with p-values ($\beta_1=0.50$ and $\beta_2=0.34$) greater than 0.05, indicating the absence of linear and quadratic trends, or equivalently, a random behaviour of the residuals with respect to the reference values. Finally, Figure 7 shows the histogram of the residuals for both calibration and cross-validation samples, note that the peak is close to zero.

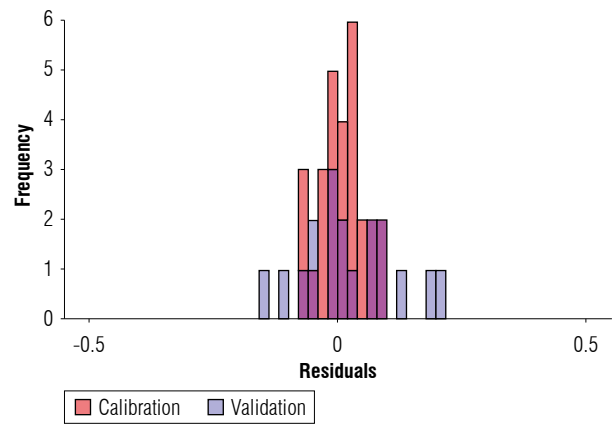


FIGURE 7. Histogram of the residuals for the selected partition. The maximum of the distribution is close to zero.

TABLE 5. Predicted cadmium concentrations by LIBS in 46 samples of cocoa beans. In parentheses, the residual corresponding to each value predicted by the model.

Sample	Cd Concentration (mg kg ⁻¹)	Sample	Cd Concentration (mg kg ⁻¹)
M01	0.19(0.129)	M24	0.02(-0.002)
M02	0.87(-0.034)	M25	0.01(0.031)
M03	0.68(0.0249)	M26	0.009(0.039)
M04	0.46(-0.153)	M27	0.05(0.003)
M05	0.29(-0.058)	M28	0.08(-0.063)
M06	0.56(0.047)	M29	0.08(-0.035)
M07	0.75(0.206)	M30	0.014(0.019)
M08	0.61(-0.069)	M31	0.04(-0.004)
M09	0.52(0.088)	M32	0.31(-0.008)
M10	0.0083(0.183)	M33	-0.01(0.059)
M11	0.22(0.094)	M34	0.10(-0.049)
M12	0.20(0.018)	M35	0.09(-0.069)
M13	0.08(0.096)	M36	0.079(-0.004)
M14	0.18(-0.011)	M37	0.09(-0.078)
M15	0.07(0.096)	M38	0.0716(-0.054)
M16	0.16(0.006)	M39	-0.0164(0.022)
M17	0.10(0.065)	M40	0.17(-0.100)
M18	0.05(0.076)	M41	0.14(0.065)
M19	0.15(-0.003)	M42	0.14(0.018)
M20	0.25(0.039)	M43	0.12(0.037)
M21	0.04(0.009)	M44	0.21(0.074)
M22	0.03(0.025)	M45	0.97(-0.012)
M23	0.039(-0.007)	M46	0.08(-0.026)

Conclusions

In this research, a PLS-1 predictive model was developed to determine the concentration of Cd in cocoa beans from a LIBS spectra. The performance metrics (RER) allow classifying the model as suitable for screening, mainly in the range of 0.2 to 1 mg kg⁻¹. The Monte Carlo cross-validation strategy allowed the selection of a sample partition that yielded a model with adequate correlation, a linear behaviour, which was evident from the diagnostic figures of merit (R², residuals and standard errors) and inference tests (F-test). Moreover, it was possible to implement the SLS transformation for the analysis of cocoa beans using LIBS, which significantly improved the signal-to-noise ratio of the spectrum allowing the development of this predictive model with relative experimental simplicity and few chemical reagents, reducing the time compared to other methodologies. The enhancement factor in the predicted concentrations, before and after SLS transformation, was estimated to be around two orders.

Acknowledgments

This research was financially supported by the Vicerrectoría de Investigación y Extensión - Universidad Industrial de Santander, Colombia (VIE-UIS), Internal Project: code 2812.

Conflict of interest statement

The authors declare that there is no conflict of interests regarding the publication of this article.

Author's contributions

RCH conceptualized and devised the methodology. SLHC collected data and contributed to the methodology and formal analysis. JEGB and EMO carried out the formal analysis. All authors contributed to the writing and reviewing of the final version of the manuscript.

Literature cited

- Aguirre-Forero, S. E., Piraneque-Gambasica, N. V., & Vásquez-Polo, J. R. (2020). Heavy metals content in soils and cocoa tissues in Magdalena department Colombia: emphasis in cadmium. *Entramado*, 16(2), 298–310. <https://doi.org/10.18041/1900-3803/entramado.2.6753>
- Andrade-Garda, J. M. (2009). *Basic chemometric techniques in atomic spectroscopy*. RSC.
- Argüello, D., Chavez, E., Laurysen, F., Vanderschueren, R., Smolders, E., & Montalvo, D. (2019). Soil properties and agronomic factors affecting cadmium concentrations in cacao beans: A nationwide survey in Ecuador. *Science of the Total Environment*, 649, 120–127. <https://doi.org/10.1016/j.scitotenv.2018.08.292>
- ASTM E1655-05. (2012). *Standard practices for infrared multivariate quantitative analysis*. ASTM International. <https://doi.org/10.1520/E1655-17>
- Bertoldi, D., Barbero, A., Camin, F., Caligiani, A., & Larcher, R. (2016). Multielemental fingerprinting and geographic traceability of *Theobroma cacao* beans and cocoa products. *Food Control*, 65, 46–53. <https://doi.org/10.1016/j.foodcont.2016.01.013>
- Bravo, D., Leon-Moreno, C., Martínez, C. A., Varón-Ramírez, V. M., Araujo-Carrillo, G. A., Vargas, R., Quiroga-Mateus, R., Zamora, A., & Gutierrez Rodríguez, E. A. (2021). The first national survey of cadmium in cacao farm soil in Colombia. *Agronomy*, 11(4), Article 761. <https://doi.org/10.3390/agronomy11040761>
- Chavez, E., He, Z. L., Stoffella, P. J., Mylavarapu, R. S., Li, Y. C., Moyano, B., & Baligar, V. C. (2015). Concentration of cadmium in cacao beans and its relationship with soil cadmium in southern Ecuador. *Science of the Total Environment*, 533, 205–214. <https://doi.org/10.1016/j.scitotenv.2015.06.106>
- Checa, K., Gamarra, M., Soto, J., Ipanaqué, W., & La Rosa, G. (2019). Preliminary study of the relation between the content of cadmium and the hyperspectral signature of organic cocoa beans. *Proceedings of the Conference on Electrical, Electronics Engineering, Information and Communication*

- Technologies (CHILECON)*, 1–7. <https://doi.org/10.1109/CHILECON47746.2019.8987991>
- De Jong, S. (1993). SIMPLS: An alternative approach to partial least-squares regression. *Chemometric and Intelligent Laboratory Systems*, 18(3), 251–263. [https://doi.org/10.1016/0169-7439\(93\)85002-X](https://doi.org/10.1016/0169-7439(93)85002-X)
- Escobar, S., Santander, M., Useche, P., Contreras, C., & Rodríguez, J. (2020). Aligning strategic objectives with research and development activities in a soft commodity sector: A technological plan for Colombian cocoa producers. *Agriculture*, 10(5), Article 141. <https://doi.org/10.3390/agriculture10050141>
- Escobar, S., Santander, M., Zuluaga, M., Chacón, I., Rodríguez, J., & Vaillant, F. (2021). Fine cocoa beans production: Tracking aroma precursors through a comprehensive analysis of flavor attributes formation. *Food Chemistry*, 365, Article 130627. <https://doi.org/10.1016/j.foodchem.2021.130627>
- Faber, N. M., & Ferré, F. J. (2008). On the numerical stability of two widely used PLS algorithms. *Journal of Chemometrics*, 22(2), 101–105. <https://doi.org/10.1002/cem.1112>
- Fedecacao. (2021). *Año cacaoero 2020-2021, el de mayor producción de cacao en la historia de Colombia*. News. <https://www.fedecacao.com.co/post/año-cacaoero-2020-2021-el-de-mayor-producción-de-cacao-en-la-historia-de-colombia>
- Filgueiras, P. R., Alves, J. C. L., Sad, C. M. S., Castro, E. V. R., Dias, J. C. M., & Poppi, R. J. (2014). Evaluation of trends in residuals of multivariate calibration models by permutation test. *Chemometrics and Intelligent Laboratory Systems*, 133, 33–41. <https://doi.org/10.1016/j.chemolab.2014.02.002>
- Gramlich, A., Tandy, S., Andres, C., Chincheros Paniagua, J., Armengot, L., Schneider, M., & Schulin, R. (2017). Cadmium uptake by cocoa trees in agroforestry and monoculture systems under conventional and organic management. *Science of the Total Environment*, 580, 677–686. <https://doi.org/10.1016/j.scitotenv.2016.12.014>
- Järup, L., & Åkesson, A. (2009). Current status of cadmium as an environmental health problem. *Toxicology and Applied Pharmacology*, 238(3), 201–208. <https://doi.org/10.1016/j.taap.2009.04.020>
- Jorhem, L., & Engman, J. (2000). Determination of lead, cadmium, zinc, copper, and iron in foods by atomic absorption spectrometry after microwave digestion: NMKL1 collaborative study. *Journal of AOAC International*, 83(5), 1189–1203.
- Menegatti, C. R., Nicolodelli, G., Senesi, G. S., da Silva, O. A., Filho, H. J. I., Villas-Boas, P. R., Marangoni, B. S., & Milori, D. M. B. P. (2019). Evaluation of LIBS under controlled atmosphere to quantify cadmium at low concentration in landfill leachates. *Applied Physics B*, 125, Article 74. <https://doi.org/10.1007/s00340-019-7189-9>
- Nicolodelli, G., Cabral, J., Menegatti, C. R., Marangoni, B., & Senesi, G. S. (2019). Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives: An overview of developments in the last decade (2010–2019). Part I. Soils and fertilizers. *TrAC - Trends in Analytical Chemistry*, 115, 70–82. <https://doi.org/10.1016/j.trac.2019.03.032>
- Niño, A. R., Ramírez, C. X., Hernández, R. C., Picón, H., Guerrero, J. E., & Mejía-Ospino, E. (2019). FTIR-ATR predictive model for determination of asphaltene solubility class index (ASCI) based on partial least-squares regression (PLS-R). *Energy and Fuels*, 33(12), 12213–12218. <https://doi.org/10.1021/acs.energyfuels.9b02829>
- Oliveira, A. P. F., Milani, R. F., Efraim, P., Morgano, M. A., & Tfouni, S. A. V. (2021). Cd and Pb in cocoa beans: Occurrence and effects of chocolate processing. *Food Control*, 119, Article 107455. <https://doi.org/10.1016/j.foodcont.2020.107455>
- Otto, M. (2007). *Chemometrics: Statistics and computer application in analytical chemistry* (2nd ed.). Weinheim, Wiley-VCH.
- Rodríguez Albarracín, H. S., Darghan Contreras, A. E., & Henao, M. C. (2019). Spatial regression modeling of soils with high cadmium content in a cocoa producing area of Central Colombia. *Geoderma Regional*, 16, Article e00214. <https://doi.org/10.1016/j.geodrs.2019.e00214>
- Rodríguez Giraldo, Y., Rodríguez Sánchez, S., Torres, L. G., Montenegro, A. C., & Pichimata, M. A. (2022). Development of validation methods to determine cadmium in cocoa almond from the beans by ICP-MS and ICP-OES. *Talanta Open*, 5, Article 100078. <https://doi.org/10.1016/j.talo.2021.100078>
- Satarug, S. (2018). Dietary cadmium intake and its effects on kidneys. *Toxics*, 6(1), Article 15. <https://doi.org/10.3390/toxics6010015>
- Senesi, G. S., Cabral, J., Menegatti, C. R., Marangoni, B., & Nicolodelli, G. (2019). Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives: An overview of developments in the last decade (2010–2019). Part II. Crop plants and their food derivatives. *TrAC - Trends in Analytical Chemistry*, 118, 453–469. <https://doi.org/10.1016/j.trac.2019.05.052>
- Sezer, B., Bilge, G., & Boyaci, I. H. (2017). Capabilities and limitations of LIBS in food analysis. *TrAC - Trends in Analytical Chemistry*, 97, 345–353. <https://doi.org/10.1016/j.trac.2017.10.003>
- Shen, T., Kong, W., Liu, F., Chen, Z., Yao, J., Wang, W., Peng, J., Chen, H., & He, Y. (2018). Rapid determination of cadmium contamination in lettuce using laser-induced breakdown spectroscopy. *Molecules*, 23(11), Article 2930. <https://doi.org/10.3390/molecules23112930>
- Vanderschueren, R., De Mesmaeker, V., Mounicou, S., Isaure, M. P., Doelsch, E., Montalvo, D., Delcour, J. A., Chavez, E., & Smolders, E. (2020). The impact of fermentation on the distribution of cadmium in cacao beans. *Food Research International*, 127, Article 108743. <https://doi.org/10.1016/j.foodres.2019.108743>
- Wang, W., Kong, W., Shen, T., Man, Z., Zhu, W., He, Y., & Liu, F. (2021). Quantitative analysis of cadmium in rice roots based on LIBS and chemometrics methods. *Environmental Sciences Europe*, 33(1), Article 37. <https://doi.org/10.1186/s12302-021-00480-4>
- Xu, Q. S., & Liang, Y. Z. (2001). Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, 56(1), 1–11. [https://doi.org/10.1016/S0169-7439\(00\)00122-2](https://doi.org/10.1016/S0169-7439(00)00122-2)
- Yang, P., Zhou, R., Zhang, W., Yi, R., Tang, S., Guo, L., Hao, Z., Li, X., Lu, Y., & Zeng, X. (2019). High-sensitivity determination of cadmium and lead in rice using laser-induced breakdown spectroscopy. *Food Chemistry*, 272, 323–328. <https://doi.org/10.1016/j.foodchem.2018.07.214>
- Yang, P., Zhu, Y., Yang, X., Li, J., Tang, S., Hao, Z., Guo, L., Li, X., Zeng, X., & Lu, Y. (2018). Evaluation of sample preparation

- methods for rice geographic origin classification using laser-induced breakdown spectroscopy. *Journal of Cereal Science*, 80, 111–118. <https://doi.org/10.1016/j.jcs.2018.01.007>
- Yao, M., Yang, H., Huang, L., Chen, T., Rao, G., & Liu, M. (2017). Detection of heavy metal Cd in polluted fresh leafy vegetables by laser-induced breakdown spectroscopy. *Applied Optics*, 56(14), 4070–4075. <https://doi.org/10.1364/ao.56.004070>
- Zhao, X., Zhao, C., Du, X., & Dong, D. (2019). Detecting and mapping harmful chemicals in fruit and vegetables using nanoparticle-enhanced laser-induced breakdown spectroscopy. *Scientific Reports*, 9(1), Article 906. <https://doi.org/10.1038/s41598-018-37556-w>